In [108…

```
print('For my second project I will be doing a Sleep Vs Health data to meaure the impact o
print('')
print('HERE IS MY CITATION: Tharmalingam, Laksika. “Sleep Health and Lifestyle Dataset.” K
print('')
print('HERE IS THE LINK TO SOURCE: https://www.kaggle.com/datasets/uom190346a/sleep-health
```

```
For my second project I will be doing a Sleep Vs Health data to meaure the imapct of sleep
on ages and different types of occupations. I chose this dataset because I found the facto
rs playing into the quality of sleep fascinating. I tried to make this as my Project 1 in
Rstudio, but I had much difficulity, so I really wanted to make this my pyhton project!

HERE IS MY CITATION: Tharmalingam, Laksika. “Sleep Health and Lifestyle Dataset.” Kaggle,
18 Sept. 2023, www.kaggle.com/datasets/uom190346a/sleep-health-and-lifestyle-dataset

HERE IS THE LINK TO SOURCE: https://www.kaggle.com/datasets/uom190346a/sleep-health-and-li
festyle-dataset
```

In [1]:

```
print('Objective: To systematically analyze and compare sleep patterns, including sleep du
```

```
Objective: To systematically analyze and compare sleep patterns, including sleep duration
and quality, across diverse occupational groups. The goal is to discern potential correlat
ions or variations in sleep behaviors among different professions, aiming to uncover insig
hts about how occupational factors influence sleep health
```

In [89]:

```
import pandas as pd

file_path = 'Desktop/Sleep/Sleep.csv'

# Load the CSV file into a DataFrame
data = pd.read_csv(file_path)

# Display the first few rows
print(data.head())
```

```
   Person ID Gender  Age            Occupation  Sleep Duration  \
0          1   Male   27     Software Engineer             6.1
1          2   Male   28                Doctor             6.2
2          3   Male   28                Doctor             6.2
3          4   Male   28  Sales Representative             5.9
4          5   Male   28  Sales Representative             5.9

   Quality of Sleep  Physical Activity Level  Stress Level BMI Category  \
0                 6                       42             6   Overweight
1                 6                       60             8       Normal
2                 6                       60             8       Normal
3                 4                       30             8        Obese
4                 4                       30             8        Obese

  Blood Pressure  Heart Rate  Daily Steps Sleep Disorder
0         126/83          77         4200            None
1         125/80          75        10000            None
2         125/80          75        10000            None
3         140/90          85         3000    Sleep Apnea
4         140/90          85         3000    Sleep Apnea
```

In [77]:

```
print('Here we are importing the dataset and important libraries needed to do our graphs a
```

```
Here we are importing the dataset and important libraries needed to do our graphs and calc
ulations. We are also doing Exploratory Data Analysis, which is basically displaying the f
irst few rows of the dataset to get a quick view of its structure.
```

In [6]:

```
# Get summary statistics for numerical columns
print(data.describe())

# Check the data types of columns
print(data.dtypes)

# Count unique values in categorical columns
print(data['Gender'].value_counts())
print(data['Occupation'].value_counts())
# ...

# Check for missing values
print(data.isnull().sum())
```

```
        Person ID         Age  Sleep Duration  Quality of Sleep  \
count  374.000000  374.000000      374.000000        374.000000
mean   187.500000   42.184492        7.132086          7.312834
std    108.108742    8.673133        0.795657          1.196956
min      1.000000   27.000000        5.800000          4.000000
25%     94.250000   35.250000        6.400000          6.000000
50%    187.500000   43.000000        7.200000          7.000000
75%    280.750000   50.000000        7.800000          8.000000
max    374.000000   59.000000        8.500000          9.000000

       Physical Activity Level  Stress Level  Heart Rate   Daily Steps
count               374.000000    374.000000  374.000000    374.000000
mean                 59.171123      5.385027   70.165775   6816.844920
std                  20.830804      1.774526    4.135676   1617.915679
min                  30.000000      3.000000   65.000000   3000.000000
25%                  45.000000      4.000000   68.000000   5600.000000
50%                  60.000000      5.000000   70.000000   7000.000000
75%                  75.000000      7.000000   72.000000   8000.000000
max                  90.000000      8.000000   86.000000  10000.000000
Person ID                  int64
Gender                    object
Age                        int64
Occupation                object
Sleep Duration           float64
Quality of Sleep           int64
Physical Activity Level    int64
Stress Level               int64
BMI Category              object
Blood Pressure            object
Heart Rate                 int64
Daily Steps                int64
Sleep Disorder            object
dtype: object
Male      189
Female    185
Name: Gender, dtype: int64
Nurse                   73
Doctor                  71
Engineer                63
Lawyer                  47
Teacher                 40
Accountant              37
Salesperson             32
Software Engineer        4
Scientist                4
Sales Representative     2
Manager                  1
Name: Occupation, dtype: int64
Person ID                  0
```

```
Gender                     0
Age                        0
Occupation                 0
Sleep Duration             0
Quality of Sleep           0
Physical Activity Level    0
Stress Level               0
BMI Category               0
Blood Pressure             0
Heart Rate                 0
Daily Steps                0
Sleep Disorder             0
dtype: int64
```

In [83]:

```
print('Here is just a summerization of all data within our Sleep and Health Dataset with t
```

```
Here is just a summerization of all data within our Sleep and Health Dataset with the coun
ts. These are numeric values that will come in handy when calculating averages and p-value
s
```

In [113…

```
import matplotlib.pyplot as plt

# Histogram of Age
plt.hist(data['Occupation'], bins=20, color='lightcoral')
plt.xlabel('Occupation')
plt.ylabel('Frequency')
plt.title('Distribution of Occupation')
plt.xticks(rotation=55)

plt.show()
```

In [79]:

```
print('The histogram visually displays how ages are distributed across the Sleep Vs. Healt
```

```
The histogram visually displays how ages are distributed across the Sleep Vs. Health datas
et. This historgram is cruical in our analysis because it helps us to understanding the sp
read of different age groups. It reveals whether the ages are evenly distributed or skewed
either to the right or left.
```

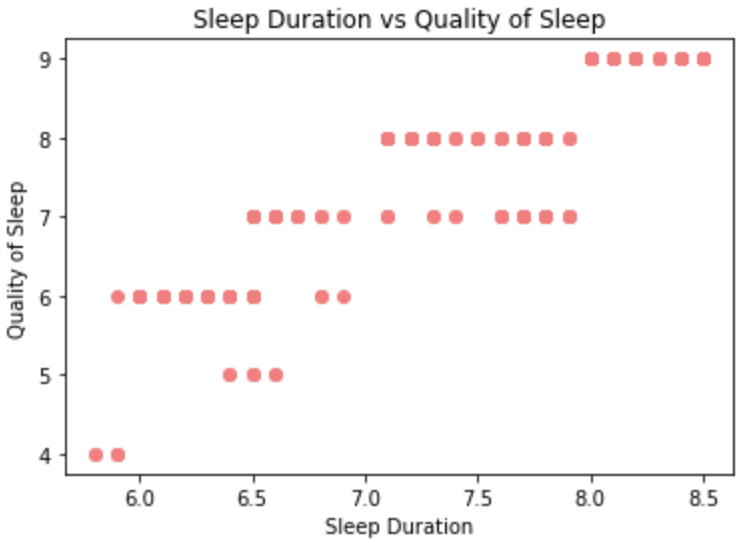
In [74]:
```
# Handling missing values (if any)
data.dropna(inplace=True)  # Drop rows with missing values

# Removing duplicates
data.drop_duplicates(inplace=True)
```

In [ ]:
```
```

In [72]:
```
import matplotlib.pyplot as plt

# Scatter plot of Sleep Duration vs. Quality of Sleep
plt.scatter(data['Sleep Duration'], data['Quality of Sleep'], color='lightcoral')
plt.xlabel('Sleep Duration')
plt.ylabel('Quality of Sleep')
plt.title('Sleep Duration vs Quality of Sleep')
plt.show()
```



In [13]:
```
relevant_columns = ['Occupation', 'Sleep Duration', 'Quality of Sleep']
data = data[relevant_columns]

occupation_sleep = data.groupby('Occupation').agg({'Sleep Duration': 'mean', 'Quality of S

print(occupation_sleep)
```

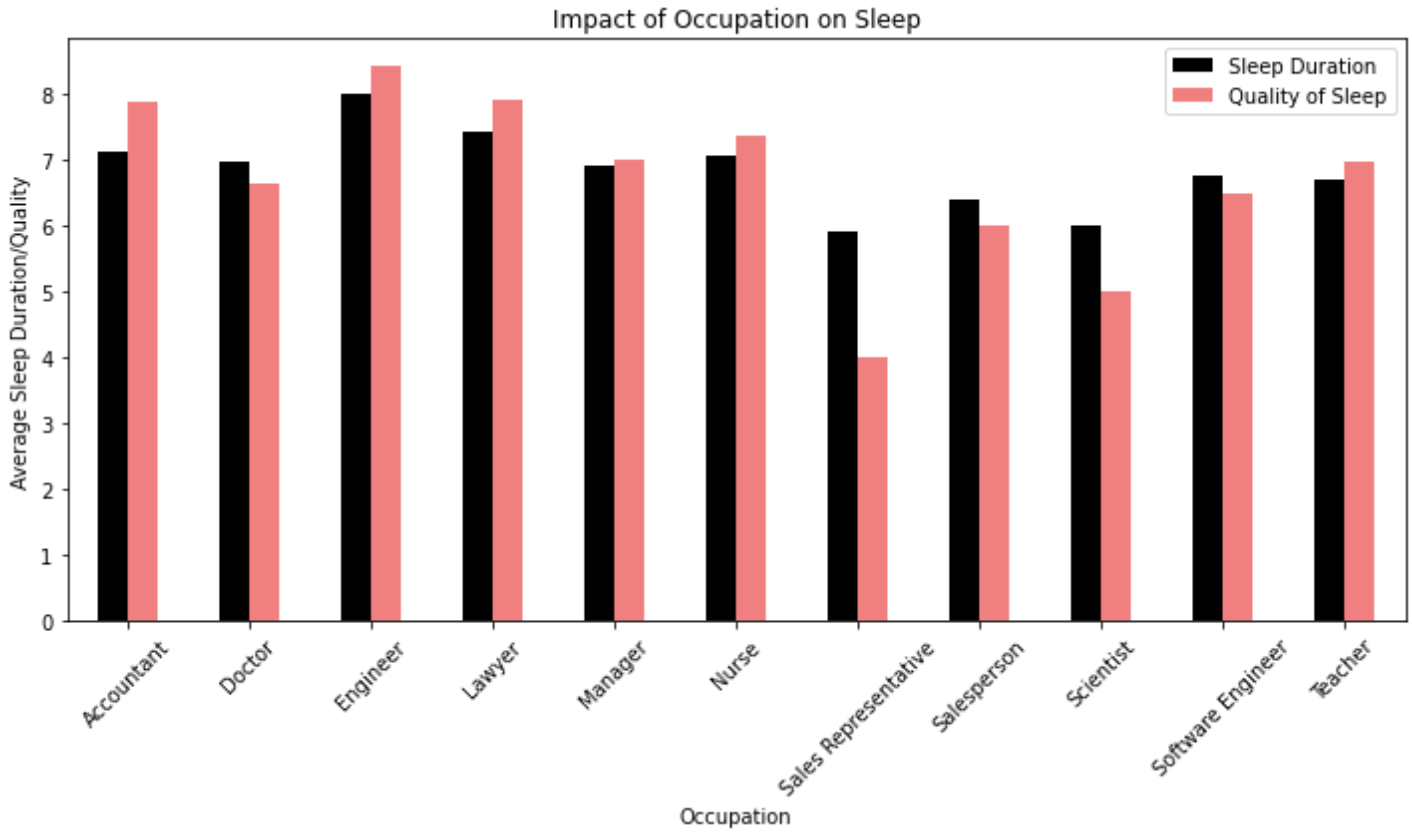
Occupation	Sleep Duration	Quality of Sleep
Accountant	7.113514	7.891892
Doctor	6.970423	6.647887
Engineer	7.987302	8.412698
Lawyer	7.410638	7.893617
Manager	6.900000	7.000000
Nurse	7.063014	7.369863
Sales Representative	5.900000	4.000000
Salesperson	6.403125	6.000000
Scientist	6.000000	5.000000
Software Engineer	6.750000	6.500000
Teacher	6.690000	6.975000

In [85]:
```
print('The data above provides a summary of average sleep duration and quality for each oc
```

The data above provides a summary of average sleep duration and quality for each occupation. From this data on Sleep Duration and Sleep Quality, we can easily and visably compare and analyze the differences between occupations and their effects on sleep. This is an important step in our analysis because it can begin to reveal patterns or trends indicating how different occupations correlate with sleep duration and quality.

In [47]:
```
 # Custom colors for bars
custom_colors = ['black', 'lightcoral']

# Plotting the bar chart
occupation_sleep.plot(kind='bar', figsize=(10, 6), color=custom_colors)
plt.title('Impact of Occupation on Sleep')
plt.xlabel('Occupation')
plt.ylabel('Average Sleep Duration/Quality')
plt.xticks(rotation=45)
plt.legend(loc='upper right')
plt.tight_layout()
plt.show()
```



In [82]:
```
print('This bar chart provides a visual comparison of the average sleep duration and quali
```

This bar chart provides a visual comparison of the average sleep duration and quality across different occupations. Looking at the bar chart, we can compare metrics like Average Sleep Duration among occupations represented by individual bars and their colors. The use of custom colors helps in differentiating between sleep duration and quality within each occupation category, enhancing readability and interpretation. This visualization helps in identifying potential trends or differences in sleep patterns among various occupations within our Sleep Health dataset.

In [88]:
```
# ANOVA for Sleep Duration Across Occupations
import scipy.stats as stats

# Extract data for ANOVA
occupation_groups = [data[data['Occupation'] == occupation]['Sleep Duration'] for occupati
```

```
# Perform ANOVA test
anova_result = stats.f_oneway(*occupation_groups)
print("ANOVA p-value:", anova_result.pvalue)

print("Since the p-value is extremely smaller than 0.05, the p-value strongly supports the
```

```
ANOVA p-value: 2.9117522724389375e-30
Since the p-value is extremely smaller than 0.05, the p-value strongly supports the reject
ion of the null hypothesis. Therefore, it indicates that there are statistically significa
nt differences in sleep durations among the various occupations
```

In [73]:

```
from statsmodels.stats.multicomp import pairwise_tukeyhsd

# Run Tukey's HSD test, printt the summary
pairwise_tukeyhsd(data['Sleep Duration'], data['Occupation']).summary()
```

Out[73]:

Multiple Comparison of Means - Tukey HSD, FWER=0.05

group1	group2	meandiff	p-adj	lower	upper	reject
Accountant	Doctor	-0.1431	0.9	-0.5661	0.2799	False
Accountant	Engineer	0.8738	0.001	0.4417	1.3059	True
Accountant	Lawyer	0.2971	0.5676	-0.1614	0.7557	False
Accountant	Manager	-0.2135	0.9	-2.3279	1.9008	False
Accountant	Nurse	-0.0505	0.9	-0.4715	0.3705	False
Accountant	Sales Representative	-1.2135	0.255	-2.7281	0.3011	False
Accountant	Salesperson	-0.7104	0.001	-1.214	-0.2067	True
Accountant	Scientist	-1.1135	0.0437	-2.2116	-0.0154	True
Accountant	Software Engineer	-0.3635	0.9	-1.4616	0.7346	False
Accountant	Teacher	-0.4235	0.1321	-0.8994	0.0524	False
Doctor	Engineer	1.0169	0.001	0.6558	1.378	True
Doctor	Lawyer	0.4402	0.014	0.0479	0.8325	True
Doctor	Manager	-0.0704	0.9	-2.1714	2.0306	False
Doctor	Nurse	0.0926	0.9	-0.2552	0.4403	False
Doctor	Sales Representative	-1.0704	0.4274	-2.5663	0.4255	False
Doctor	Salesperson	-0.5673	0.0021	-1.0115	-0.1231	True
Doctor	Scientist	-0.9704	0.1176	-2.0426	0.1017	False
Doctor	Software Engineer	-0.2204	0.9	-1.2926	0.8517	False
Doctor	Teacher	-0.2804	0.5039	-0.6929	0.132	False
Engineer	Lawyer	-0.5767	0.001	-0.9788	-0.1745	True
Engineer	Manager	-1.0873	0.829	-3.1901	1.0155	False
Engineer	Nurse	-0.9243	0.001	-1.2831	-0.5655	True
Engineer	Sales Representative	-2.0873	0.001	-3.5858	-0.5888	True
Engineer	Salesperson	-1.5842	0.001	-2.0371	-1.1313	True
Engineer	Scientist	-1.9873	0.001	-3.0631	-0.9115	True
Engineer	Software Engineer	-1.2373	0.0102	-2.3131	-0.1615	True
Engineer	Teacher	-1.2973	0.001	-1.7191	-0.8755	True
Lawyer	Manager	-0.5106	0.9	-2.6191	1.5978	False

Lawyer	Nurse	-0.3476	0.1311	-0.7378	0.0426	False
Lawyer	Sales Representative	-1.5106	0.0487	-3.017	-0.0043	True
Lawyer	Salesperson	-1.0075	0.001	-1.4857	-0.5294	True
Lawyer	Scientist	-1.4106	0.0016	-2.4973	-0.324	True
Lawyer	Software Engineer	-0.6606	0.6475	-1.7473	0.426	False
Lawyer	Teacher	-0.7206	0.001	-1.1695	-0.2718	True
Manager	Nurse	0.163	0.9	-1.9376	2.2636	False
Manager	Sales Representative	-1.0	0.9	-3.5552	1.5552	False
Manager	Salesperson	-0.4969	0.9	-2.6156	1.6218	False
Manager	Scientist	-0.9	0.9	-3.2326	1.4326	False
Manager	Software Engineer	-0.15	0.9	-2.4826	2.1826	False
Manager	Teacher	-0.21	0.9	-2.3223	1.9023	False
Nurse	Sales Representative	-1.163	0.297	-2.6584	0.3323	False
Nurse	Salesperson	-0.6599	0.001	-1.1022	-0.2176	True
Nurse	Scientist	-1.063	0.0539	-2.1344	0.0084	False
Nurse	Software Engineer	-0.313	0.9	-1.3844	0.7584	False
Nurse	Teacher	-0.373	0.1139	-0.7834	0.0374	False
Sales Representative	Salesperson	0.5031	0.9	-1.0175	2.0238	False
Sales Representative	Scientist	0.1	0.9	-1.7068	1.9068	False
Sales Representative	Software Engineer	0.85	0.9	-0.9568	2.6568	False
Sales Representative	Teacher	0.79	0.818	-0.7217	2.3017	False
Salesperson	Scientist	-0.4031	0.9	-1.5096	0.7033	False
Salesperson	Software Engineer	0.3469	0.9	-0.7596	1.4533	False
Salesperson	Teacher	0.2869	0.7038	-0.2079	0.7817	False
Scientist	Software Engineer	0.75	0.8464	-0.7253	2.2253	False
Scientist	Teacher	0.69	0.6022	-0.4041	1.7841	False
Software Engineer	Teacher	-0.06	0.9	-1.1541	1.0341	False

In [4]:
```
print("The Tukey's range test (or Tukey's Honestly Significant Difference test - Tukey HSD
```

```
The Tukey's range test (or Tukey's Honestly Significant Difference test - Tukey HSD) is a
statistical method. It is a function taken from the python Library. It is typically an Ana
lysis of Variance to determine which specific groups differ significantly from each other.
In this case, we are ultizing different occupations and their correlaion with quality of s
leep.
```

In [2]:
```
print('CONCLUSION: The statistical significance suggests that the average sleep durations
```

```
CONCLUSION: The statistical significance suggests that the average sleep durations signifi
cantly differ across the occupations included in the analysis. These findings might be val
uable for occupational health interventions or policies aimed at improving sleep patterns
within specific professions Based on these significant differences, tailored approaches or
strategies to address sleep-related issues within certain occupational groups could be rec
ommended
```

In [ ]:
```

```